Supplemental Material

I. BOUNDS ON PERTURBATION MODULARITY

The bounds on perturbation modularity depend on the norm used to measure perturbation magnitude (see Eq. (1) in the main text). Here we consider the family of ℓ_p norms, where the ℓ_p norm of vector $\boldsymbol{a} = \langle a_1, \ldots, a_N \rangle$ is defined for $p > 0$ as $\|\boldsymbol{a}\|_p = \left(\sum_{i=1}^N |a_i|^p\right)^{1/p}$.

As we show below, perturbation modularity is bounded by -1 and 1 for ℓ_1 and ℓ_2 norms. More generally, for ℓ_p norms with $p \geq 2$, perturbation modularity is bounded by $-N^{(p-2/p)}$ and $N^{(p-2/p)}$, where N is the number of variables in the original system.

Assume some ℓ_p norm is used to measure perturbation magnitude. Using the definition of perturbation modularity for initial condition $\boldsymbol{x}$ and partition π:

$$\begin{aligned} Q^t(\pi, \boldsymbol{x}) &= \mathrm{E}\big[\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon}) \cdot \boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] - \mathrm{E}\big[\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] \cdot \mathrm{E}\big[\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] \\ &\leq \mathrm{E}\big[\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon}) \cdot \boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] \\ &\leq \mathrm{E}\big[\big\|\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2 \big\|\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2\big] \end{aligned}$$

where the second line follows from the non-negativity of the coarse-grained perturbation vectors $\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})$ and $\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})$ and the third line follows from the Cauchy–Schwarz inequality.

When some ℓ_p norm is used to measure perturbation magnitude, the coarse-grained perturbation vectors themselves are unit vectors in ℓ_p. To show this, let $v_i = |\left(f^t(\boldsymbol{x} + \boldsymbol{\varepsilon}) - f^t(\boldsymbol{x})\right)_{\{i\}}|$. Then,

$$\big\|\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_p = \left(\sum_{S\in\pi} \left[m_S^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\right]^p\right)^{1/p} = \left(\sum_{S\in\pi} \left[\frac{\left(\sum_{i\in S} v_i^p\right)^{\frac{1}{p}}}{\left(\sum_{i'=1}^N v_{i'}^p\right)^{\frac{1}{p}}}\right]^p\right)^{1/p} = \left(\sum_{S\in\pi} \frac{\sum_{i\in S} v_i^p}{\sum_{i'=1}^N v_{i'}^p}\right)^{1/p} = 1$$

When $p = 2$ is used to measure perturbation magnitudes, the upper bound can be rewritten as:

$$Q^t(\pi, \boldsymbol{x}) \leq \mathrm{E}\big[\big\|\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2 \big\|\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2\big] = \mathrm{E}[1 \cdot 1] = 1$$

When $p = 1$ is used to measure perturbation magnitudes, we first note that $\|\boldsymbol{a}\|_2 \leq \|\boldsymbol{a}\|_1$ for any $\boldsymbol{a}$ and that $\big\|\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_1 = \|\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\|_1 = 1$. The upper bound can be rewritten as:

$$Q^t(\pi, \boldsymbol{x}) \leq \mathrm{E}\big[\big\|\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2 \big\|\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2\big] \leq \mathrm{E}[1 \cdot 1] = 1$$

More generally, when $p \geq 2$, Hölder's inequality [1] provides the bound $\|\boldsymbol{a}\|_2 \leq n^{\left(\frac{1}{2} - \frac{1}{p}\right)} \|\boldsymbol{a}\|_p$, where n is the number of dimensions of $\boldsymbol{a}$. Thus, $\|\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\|_2 \leq |\pi|^{\left(\frac{1}{2} - \frac{1}{p}\right)} \|\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\|_p = |\pi|^{\left(\frac{1}{2} - \frac{1}{p}\right)} \leq N^{\left(\frac{1}{2} - \frac{1}{p}\right)}$, since N is the maximum number of subsets in a partition. This gives:

$$Q^t(\pi, \boldsymbol{x}) \leq \mathrm{E}\big[\big\|\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2 \big\|\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2\big] \leq \mathrm{E}\Big[N^{\left(\frac{1}{2} - \frac{1}{p}\right)} N^{\left(\frac{1}{2} - \frac{1}{p}\right)}\Big] = N^{\left(\frac{p-2}{p}\right)}$$

To show the lower bound, we again use the definition of perturbation modularity and similar reasoning:

$$\begin{aligned} Q^t(\pi, \boldsymbol{x}) &= \mathrm{E}\big[\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon}) \cdot \boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] - \mathrm{E}\big[\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] \cdot \mathrm{E}\big[\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] \\ &\geq -\mathrm{E}\big[\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] \cdot \mathrm{E}\big[\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] \\ &\geq -\big\|\mathrm{E}\big[\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big]\big\|_2 \big\|\mathrm{E}\big[\boldsymbol{y}_\pi^t(\boldsymbol{x}, \boldsymbol{\varepsilon})\big]\big\|_2 \\ &\geq -\mathrm{E}\big[\big\|\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2\big] \, \mathrm{E}\big[\big\|\boldsymbol{y}_\pi^0(\boldsymbol{x}, \boldsymbol{\varepsilon})\big\|_2\big] \end{aligned}$$

where the last line follows from Jensen's inequality and the convexity of norms. Similar arguments as above show that $Q^t(\pi, \boldsymbol{x}) \geq -1$ for $p = 1$ and $p = 2$ and $Q^t(\pi, \boldsymbol{x}) \geq -N^{(p-2/p)}$ for $p > 2$.

In practice, we are also interested in the maximal perturbation modularity across all partitions. In fact, maximal perturbation modularity is always non-negative. This is because there is at least one partition with perturbation modularity equal to 0: the partition that contains the entire system as one subsystem. In this partition, which we call $\pi_0 = \{\{1, \ldots, N\}\}$, perturbations are entirely contained in the single subsystem and $\boldsymbol{y}_{\pi_0}^0(\boldsymbol{x}, \boldsymbol{\varepsilon}) = \boldsymbol{y}_{\pi_0}^t(\boldsymbol{x}, \boldsymbol{\varepsilon}) = \langle 1 \rangle$ for all $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$. From the definition of perturbation modularity, it can be seen that $Q^t(\pi_0, \boldsymbol{x}) = 0$ for all t.

II. PERTURBATION MODULARITY AND COMMUNITY DETECTION

An explicit connection can be made between our approach and graph-based community detection. We first rewrite and expand the definition of perturbation modularity from the main text:

$$\begin{aligned} Q^t(\pi, \boldsymbol{x}) &= \mathrm{E}\big[\boldsymbol{y}^0_\pi(\boldsymbol{x}, \boldsymbol{\varepsilon}) \cdot \boldsymbol{y}^t_\pi(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] - \mathrm{E}\big[\boldsymbol{y}^0_\pi(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] \cdot \mathrm{E}\big[\boldsymbol{y}^t_\pi(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] \\ &= \sum_{S\in\pi} \left(\mathrm{E}\big[m^0_S(\boldsymbol{x}, \boldsymbol{\varepsilon})\, m^t_S(\boldsymbol{x}, \boldsymbol{\varepsilon})\big] - \mathrm{E}\big[m^0_S(\boldsymbol{x}, \boldsymbol{\varepsilon})\big]\, \mathrm{E}\left[m^t_S(\boldsymbol{x}, \boldsymbol{\varepsilon})\right]\right) \end{aligned}$$

where the expectations are over $P(\boldsymbol{\varepsilon})$. We assume that the ℓ_1 norm is used to measure perturbation magnitude, which provides the following additive property: $m^t_S(\boldsymbol{x}, \boldsymbol{\varepsilon}) = \sum_{i\in S} m^t_{\{i\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})$, where the subscript $\{i\}$ indicates a subsystem only containing variable i. We rewrite the above equation as:

$$\begin{aligned} Q^t(\pi, \boldsymbol{x}) &= \sum_{S\in\pi} \left(\mathrm{E}\left[\sum_{i\in S} m^0_{\{i\}}(\boldsymbol{x}, \boldsymbol{\varepsilon}) \sum_{j\in S} m^t_{\{j\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\right] - \mathrm{E}\left[\sum_{i\in S} m^0_{\{i\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\right] \mathrm{E}\left[\sum_{j\in S} m^t_{\{j\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\right]\right) \\ &= \sum_{S\in\pi} \left(\sum_{i,j\in S} \mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\, m^t_{\{j\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\Big] - \sum_{i\in S} \mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\Big] \sum_{j\in S} \mathrm{E}\Big[m^t_{\{j\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\Big]\right) \end{aligned} \tag{S1}$$

where the last line comes from exchanging the order of summation and expectation.

A. Perturbation modularity is Markov Stability for diffusion dynamics

Markov stability is a method of community detection that uses the dynamics random walks over graphs [2–4]. Here, communities are defined as subgraphs that tend to trap random walkers. This method is of particular interest because it generalizes many other community detection methods, including optimization of Newman's modularity, cut size, and spectral methods [4]. Given a random walk over an N-node graph, the Markov stability of a partition π at time scale t is defined as:

$$R^t(\pi) = \sum_{S\in\pi} \Pr(\text{walker in } S \text{ at times } \mathit{0} \text{ and } t) - \Pr(\text{walker in } S \text{ at time } 0) \Pr(\text{walker in } S \text{ at time } t) \tag{S2}$$

In this framework, the optimal partition of a graph is the one that maximizes Markov stability.

Given homogenous perturbation to single variables, there is an equivalence between Markov stability and ℓ_1 perturbation modularity of diffusion dynamics. Specifically, the diffusion of random walkers on a graph can be stated in terms of a linear dynamical system $f^t(\boldsymbol{x}) = T^t \boldsymbol{x}$, where $\boldsymbol{x}(t)$ is an N-dimensional vector with $\boldsymbol{x}_i(t)$ being the density of random walkers at node i at time t, and T^t is the time scale t transition matrix (here superscript t indicates matrix power; for simplicity, we consider the discrete-time case). Assume that perturbation to variable i is indicated by $\boldsymbol{\varepsilon}^i = \mathbf{e}_i$, where $\mathbf{e}_i$ is the i^{th} N-dimensional standard basis vector (initial perturbations can be scaled by any constant without changing results). Then, $m^0_{\{j\}}(\boldsymbol{x}, \boldsymbol{\varepsilon}^i) = \frac{|\boldsymbol{\varepsilon}^i_{\{j\}}|}{\|\boldsymbol{\varepsilon}^i\|_1} = \frac{|(\mathbf{e}_i)_{\{j\}}|}{\|\mathbf{e}_i\|_1} = \delta_{i,j}$.

T^t is a transition matrix: it has positive entries and preserves the ℓ_1 norm upon matrix multiplication. Therefore:

$$m^t_{\{j\}}(\boldsymbol{x}, \boldsymbol{\varepsilon}^i) = \frac{\left|\left(f^t(\boldsymbol{x}+\boldsymbol{\varepsilon}^i) - f^t(\boldsymbol{x})\right)_{\{j\}}\right|}{\|f^t(\boldsymbol{x}+\boldsymbol{\varepsilon}^i) - f^t(\boldsymbol{x})\|_1} = \frac{\left|(T^t(\boldsymbol{x}+\boldsymbol{\varepsilon}^i) - T^t\boldsymbol{x})_{\{j\}}\right|}{\|T^t(\boldsymbol{x}+\boldsymbol{\varepsilon}^i) - T^t\boldsymbol{x}\|_1} = \frac{\left|(T^t\boldsymbol{\varepsilon}^i)_{\{j\}}\right|}{\|T^t\boldsymbol{\varepsilon}^i\|_1} = \frac{\left|(T^t\mathbf{e}_i)_{\{j\}}\right|}{\|T^t\mathbf{e}_i\|_1} = T^t_{ij}$$

The terms in Eq. (S1) can now be mapped to the terms in the Markov stability Eq. (S2):

$$\sum_{i,j\in S} \mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\, m^t_{\{j\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\Big] = \sum_{i,j\in S}\sum_{k=1}^{N} P(\boldsymbol{\varepsilon}^k)\delta_{i,k}T^t_{kj} = \sum_{i,j\in S} P(\boldsymbol{\varepsilon}^i)T^t_{ij} = \Pr(\text{walker in } S \text{ at times } \mathit{0} \text{ and } t)$$

$$\sum_{i\in S} \mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\Big] = \sum_{i\in S}\sum_{k=1}^{N} P(\boldsymbol{\varepsilon}^k)\delta_{i,k} = \sum_{i\in S} P(\boldsymbol{\varepsilon}^i) = \Pr(\text{walker in } S \text{ at time } \mathit{0})$$

$$\sum_{j\in S} \mathrm{E}\Big[m^t_{\{j\}}(\boldsymbol{x}, \boldsymbol{\varepsilon})\Big] = \sum_{j\in S}\sum_{k=1}^{N} P(\boldsymbol{\varepsilon}^k)T^t_{kj} = \Pr(\text{walker in } S \text{ at time } t)$$

Markov stability is usually defined for an equilibrium random walk, such that $p(\text{walker in } S \text{ at time } 0) = p(\text{walker in } S \text{ at time } t)$. In our framework, this is accomplished by setting $P\left(\boldsymbol{\varepsilon}^i\right)$ equal to the equilibrium probability of finding a random walker at node i.

B. Mapping to directed weighted Newman's modularity

In this section, we show that for any dynamical system, ℓ_1 perturbation modularity can be mapped to *directed weighted Newman's modularity* on a specially-constructed graph. One result of this mapping is that efficient community detection algorithms can be used to find partitions that maximize perturbation modularity.

To review, the weighted directed Newman's modularity of partition π is defined as [5]:

$$\sum_{C\in\pi}\sum_{i,j\in C}\left(w_{ij}-\frac{w_i^{out}w_j^{in}}{M}\right) \tag{S3}$$

where w_{ij} indicates edge weight from node i to node j, $w_i^{out}=\sum_j w_{ij}$ is the out-degree, $w_j^{in}=\sum_i w_{ij}$ is the in-degree, and $M=\sum_i\sum_j w_{ij}=\sum_i w_i^{out}=\sum_j w_j^{in}$ is the total strength (summations in these equations are over all nodes).

Now, for an N-dimensional dynamical system, we construct a graph with one node for each variable and edge weight from node i to j:

$$w_{ij}=\mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\,m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]$$

When the ℓ_1 norm is used, $\sum_{i=1}^N m^t_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})=1$. Thus,

$$\begin{aligned}
w_i^{out}&=\sum_{j=1}^N w_{ij}=\sum_{j=1}^N\mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\,m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]=\mathrm{E}\left[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\sum_{j=1}^N m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\right]=\mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]\\
w_j^{in}&=\sum_{i=1}^N w_{ij}=\sum_{i=1}^N\mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\,m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]=\mathrm{E}\left[\left(\sum_{i=1}^N m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\right)m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\right]=\mathrm{E}\Big[m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]\\
M&=\sum_{i=1}^N w_i^{out}=\sum_{i=1}^N\mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]=\mathrm{E}\left[\sum_{i=1}^N m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\right]=\mathrm{E}[1]=1
\end{aligned}$$

Rewriting Eq. (S1) makes the mapping to Eq. (S3) explicit:

$$\begin{aligned}
Q^t(\pi,\boldsymbol{x})&=\sum_{S\in\pi}\left(\sum_{i,j\in S}\mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\,m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]-\sum_{i\in S}\mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]\sum_{j\in S}\mathrm{E}\Big[m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]\right)\\
&=\sum_{S\in\pi}\sum_{i,j\in S}\Big(\mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\,m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]-\mathrm{E}\Big[m^0_{\{i\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]\,\mathrm{E}\Big[m^t_{\{j\}}(\boldsymbol{x},\boldsymbol{\varepsilon})\Big]\Big)\\
&=\sum_{S\in\pi}\left(\sum_{i,j\in S}w_{ij}-\frac{w_i^{out}w_j^{in}}{M}\right)
\end{aligned}$$

REFERENCES

[1] T. Tao, *Epsilon of Room, One* (American Mathematical Soc., 2010).
[2] R. Lambiotte, J.-C. Delvenne, and M. Barahona, arXiv preprint arXiv:0812.1770 (2008).
[3] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, PNAS **107**, 12755 (2010).
[4] R. Lambiotte, J.-C. Delvenne, and M. Barahona, Network Science and Engineering, IEEE Transactions on **1**, 76 (2014).
[5] A. Arenas, J. Duch, A. Fernández, and S. Gómez, New J. Phys **9**, 176 (2007).